

Ordering Positive Definite Matrices

Cyrus Mostajeran · Rodolphe Sepulchre

Received: date / Accepted: date

Abstract We introduce new partial orders on the set S_n^+ of positive definite matrices of dimension n derived from the affine-invariant geometry of S_n^+ . The orders are induced by affine-invariant cone fields, which arise naturally from a local analysis of the orders that are compatible with the homogeneous geometry of S_n^+ defined by the natural transitive action of the general linear group $GL(n)$. We then take a geometric approach to the study of monotone functions on S_n^+ and establish a number of relevant results, including an extension of the well-known Löwner-Heinz theorem derived using differential positivity with respect to affine-invariant cone fields.

Keywords Positive Definite Matrices · Partial Orders · Monotone Functions · Monotone Flows · Differential Positivity · Matrix Means

Mathematics Subject Classification (2010) 15B48 · 34C12 · 37C65 · 47H05

1 Introduction

Well-defined notions of ordering of elements of a space are of fundamental importance to many areas of applied mathematics, including the theory of monotone functions and matrix means in which orders play a defining role [17, 11, 2, 14]. Partial orders play a key part in a wide variety of applications

This work was funded by the Engineering and Physical Sciences Research Council (EPSRC) of the United Kingdom, as well as the European Research Council under the Advanced ERC Grant Agreement Switchlet n.670645.

C. Mostajeran
Department of Engineering, University of Cambridge, Cambridge, UK
E-mail: csm54@cam.ac.uk

R. Sepulchre
E-mail: r.sepulchre@eng.cam.ac.uk

across information geometry where one is interested in performing statistical analysis on sets of matrices. In such applications, the choice of order relation is often taken for granted. This choice, however, is of crucial significance since a function that is not monotone with respect to one order, may be monotone with respect to another.

We outline a geometric approach to systematically generate orders on homogeneous spaces. A homogeneous space is a manifold that admits a transitive action by a Lie group, in the sense that any two points on the manifold can be mapped onto each other by elements of a group of transformations that act on the space. The observation that cone fields induce conal orders on continuous spaces, combined with the geometry of homogeneous spaces forms the basis of the approach taken in this paper. The aim is to generate cone fields that are invariant with respect to the homogeneous geometry, thereby defining partial orders built upon the underlying symmetries of the space. A smooth cone field on a manifold is often also referred to as a causal structure. The geometry of invariant cone fields and causal structures on homogeneous spaces has been the subject of extensive studies from a Lie theoretic perspective; see [18, 12, 13], for instance. Causal structures induced by quadratic cone fields on manifolds also play a fundamental role in mathematical physics, in particular within the theory of general relativity [22].

The focus of this paper is on ordering the elements of the set of symmetric positive definite matrices S_n^+ of dimension n . Positive definite matrices arise in numerous applications, including as covariance matrices in statistics and computer vision, as variables in convex and semidefinite programming, as unknowns in fundamental problems in systems and control theory, as kernels in machine learning, and as diffusion tensors in medical imaging. The space S_n^+ forms a smooth manifold that can be viewed as a homogeneous space admitting a transitive action by the general linear group $GL(n)$, which endows the space with an affine-invariant geometry as reviewed in Section 2. In Section 3, this geometry is used to construct affine-invariant cone fields and new partial orders on S_n^+ . In Section 4, we discuss how differential positivity [9] can be used to study and characterize monotonicity on S_n^+ with respect to the invariant orders introduced in this paper. We also state and prove a generalized version of the celebrated Löwner-Heinz theorem [17, 11] of operator monotonicity theory derived using this approach. In Section 5, we consider preorder relations induced by affine-invariant and translation-invariant half-spaces on S_n^+ , and provide examples of functions and flows that preserve such structures. Finally, in Section 6, we review the notion of matrix means and establish a connection between the geometric mean and affine-invariant cone fields on S_n^+ .

2 Homogeneous geometry of S_n^+

The set S_n^+ of symmetric positive definite matrices of dimension n has the structure of a homogeneous space with a transitive $GL(n)$ -action. The transitive action of $GL(n)$ on S_n^+ is given by congruence transformations of the

form

$$\tau_A : \Sigma \mapsto A \Sigma A^T \quad \forall A \in GL(n), \forall \Sigma \in S_n^+. \quad (1)$$

Specifically, if $\Sigma_1, \Sigma_2 \in S_n^+$, then τ_A with $A = \Sigma_2^{1/2} \Sigma_1^{-1/2} \in GL(n)$ maps Σ_1 onto Σ_2 , where $\Sigma^{1/2}$ denotes the unique positive definite square root of Σ . This action is said to be *almost effective* in the sense that $\pm I$ are the only elements of $GL(n)$ that fix every $\Sigma \in S_n^+$. The isotropy group of this action at $\Sigma = I$ is precisely the orthogonal group $O(n)$, since $\tau_Q : I \mapsto QIQ^T = I$ if and only if $Q \in O(n)$. Thus, we can identify any $\Sigma \in S_n^+$ with an element of the quotient space $GL(n)/O(n)$. That is

$$S_n^+ \cong GL(n)/O(n). \quad (2)$$

The identification in (2) can also be made by noting that $\Sigma \in S_n^+$ admits a Cholesky decomposition $\Sigma = CC^T$ for some $C \in GL(n)$. The Cauchy polar decomposition of the invertible matrix C yields a unique decomposition $C = PQ$ of C into an orthogonal matrix $Q \in O(n)$ and a symmetric positive definite matrix $P \in S_n^+$. Now note that if Σ has Cholesky decomposition $\Sigma = CC^T$ and C has a Cauchy polar decomposition $C = PQ$, then $\Sigma = PQQ^T P = P^2$. That is, Σ is invariant with respect to the orthogonal part Q of the polar decomposition. Therefore, we can identify any $\Sigma \in S_n^+$ with the equivalence class $[\Sigma^{1/2}] = \Sigma^{1/2} \cdot O(n)$ in the quotient space $GL(n)/O(n)$.

Recall that the Lie algebra $\mathfrak{gl}(n)$ of $GL(n)$ consists of the set $\mathbb{R}^{n \times n}$ of all real $n \times n$ matrices equipped with the Lie bracket $[X, Y] = XY - YX$, while the Lie algebra of $O(n)$ is $\mathfrak{o}(n) = \{X \in \mathbb{R}^{n \times n} : X^T = -X\}$. Since any matrix $X \in \mathbb{R}^{n \times n}$ has a unique decomposition $X = \frac{1}{2}(X - X^T) + \frac{1}{2}(X + X^T)$, as a sum of an antisymmetric part and a symmetric part, we have $\mathfrak{gl}(n) = \mathfrak{o}(n) \oplus \mathfrak{m}$, where $\mathfrak{m} = \{X \in \mathbb{R}^{n \times n} : X^T = X\}$. Furthermore, since $\text{Ad}_Q(S) = QSQ^{-1} = QSQ^T$ is a symmetric matrix for each $S \in \mathfrak{m}$, we have

$$\text{Ad}_{O(n)} \mathfrak{m} \subseteq \mathfrak{m}, \quad (3)$$

which shows that $S_n^+ = GL(n)/O(n)$ is in fact a reductive homogeneous space with reductive decomposition $\mathfrak{gl}(n) = \mathfrak{o}(n) \oplus \mathfrak{m}$. Also, note that since $(XY - YX)^T = Y^T X^T - X^T Y^T$, we have $[\mathfrak{o}(n), \mathfrak{o}(n)] \subseteq \mathfrak{o}(n)$, $[\mathfrak{m}, \mathfrak{m}] \subseteq \mathfrak{o}(n)$, and $[\mathfrak{o}(n), \mathfrak{m}] \subseteq \mathfrak{m}$. The tangent space $T_o S_n^+$ of S_n^+ at the base-point $o = [I] = I \cdot O(n)$ is identified with \mathfrak{m} . For each $\Sigma \in S_n^+$, the action $\tau_{\Sigma^{1/2}} : S_n^+ \rightarrow S_n^+$ induces the vector space isomorphism $d\tau_{\Sigma^{1/2}}|_I : T_I S_n^+ \rightarrow T_\Sigma S_n^+$ given by

$$d\tau_{\Sigma^{1/2}}|_I X = \Sigma^{1/2} X \Sigma^{1/2}, \quad \forall X \in \mathfrak{m}. \quad (4)$$

The map (4) can be used to extend structures defined in $T_o S_n^+$ to structures defined on the tangent bundle TS_n^+ through affine-invariance, provided that the structures in $T_o S_n^+$ are $\text{Ad}_{O(n)}$ -invariant. The $\text{Ad}_{O(n)}$ -invariance is required to ensure that the extension to TS_n^+ is unique and thus well-defined. For instance, any homogeneous Riemannian metric on $S_n^+ \cong GL(n)/O(n)$ is determined by an $\text{Ad}_{O(n)}$ -invariant inner product on \mathfrak{m} . Any such inner product induces a norm that is rotationally invariant and so can only depend on the scalar invariants $\text{tr}(X^k)$ where $k \geq 1$ and $X \in \mathfrak{m}$. Moreover, as the inner product is a

quadratic function, $\|X\|^2$ must be a linear combination of $(\text{tr}(X))^2$ and $\text{tr}(X^2)$. Thus, any $\text{Ad}_{O(n)}$ -invariant inner product on \mathfrak{m} must be a scalar multiple of

$$\langle X, Y \rangle_{\mathfrak{m}} = \text{tr}(XY) + \mu \text{tr}(X) \text{tr}(Y), \quad (5)$$

where μ is a scalar parameter with $\mu > -1/n$ to ensure positive-definiteness [21]. Therefore, the corresponding affine-invariant Riemannian metrics are generated by (4) and given by

$$\begin{aligned} \langle X, Y \rangle_{\Sigma} &= \langle \Sigma^{-1/2} X \Sigma^{-1/2}, \Sigma^{-1/2} Y \Sigma^{-1/2} \rangle_{\mathfrak{m}} \\ &= \text{tr}(\Sigma^{-1} X \Sigma^{-1} Y) + \mu \text{tr}(\Sigma^{-1} X) \text{tr}(\Sigma^{-1} Y), \end{aligned} \quad (6)$$

for $\Sigma \in S_n^+$ and $X, Y \in T_{\Sigma} S_n^+$. In the case $\mu = 0$, (6) yields the most commonly used ‘natural’ Riemannian metric on S_n^+ , which corresponds to the Fisher information metric for the multivariate normal distribution [8, 23], and has been widely used in applications such as tensor computing in medical imaging [4].

3 Affine-invariant orders

3.1 Affine-invariant cone fields

A cone field \mathcal{K} on S_n^+ smoothly assigns a cone $\mathcal{K}(\Sigma) \subset T_{\Sigma} S_n^+$ to each point $\Sigma \in S_n^+$. In this paper, we consider a cone to be a solid and pointed subset of a vector space that is closed under linear combinations with positive coefficients. We say that \mathcal{K} is affine-invariant or homogeneous with respect to the quotient geometry $S_n^+ \cong GL(n)/O(n)$ if

$$d\tau_A|_{\Sigma} \mathcal{K}(\Sigma) = \mathcal{K}(\tau_A(\Sigma)), \quad (7)$$

for all $\Sigma \in S_n^+$ and $A \in GL(n)$. The procedure we will use for constructing affine-invariant cone fields on S_n^+ is similar to the approach taken for generating the affine-invariant Riemannian metrics in Section 2. We begin by defining a cone $\mathcal{K}(I)$ at I that is $\text{Ad}_{O(n)}$ -invariant:

$$X \in \mathcal{K}(I) \iff \text{Ad}_Q X = d\tau_Q|_I X = QXQ^T \in \mathcal{K}(I), \forall Q \in O(n). \quad (8)$$

Using such a cone, we generate a cone field via

$$\mathcal{K}(\Sigma) = d\tau_{\Sigma^{1/2}}|_I \mathcal{K}(I) = \{X \in T_{\Sigma} S_n^+ : \Sigma^{-1/2} X \Sigma^{-1/2} \in \mathcal{K}(I)\}. \quad (9)$$

The $\text{Ad}_{O(n)}$ -invariance condition (8) is satisfied if $\mathcal{K}(I)$ has a spectral characterization; that is, we can check to see if any given $X \in T_I S_n^+ \cong \mathfrak{m}$ lies in $\mathcal{K}(I)$ using only properties of X that are characterized by its spectrum. This observation leads to the following result.

Proposition 1 A cone $\mathcal{K}(I) \in T_I S_n^+$ is $\text{Ad}_{O(n)}$ -invariant if and only if there exists a cone $\mathcal{K}_\Lambda \subset \mathbb{R}^n$ that satisfies

$$\lambda \in \mathcal{K}_\Lambda \iff P\lambda \in \mathcal{K}_\Lambda, \quad (10)$$

for all permutation matrices $P \in \mathbb{R}^{n \times n}$, such that $X \in \mathcal{K}(I)$ whenever $\lambda_X \in \mathcal{K}_\Lambda$, where $\lambda_X = (\lambda_i(X))$ is a vector consisting of the n real eigenvalues of the symmetric matrix X .

For instance, $\text{tr}(X)$ and $\text{tr}(X^2)$ are both functions of X that are spectrally characterized and indeed $\text{Ad}_{O(n)}$ -invariant. Quadratic $\text{Ad}_{O(n)}$ -invariant cones are defined by inequalities on suitable linear combinations of $(\text{tr}(X))^2$ and $\text{tr}(X^2)$.

Proposition 2 For any choice of parameter $\mu \in (0, n)$, the set

$$\mathcal{K}(I) = \{X \in T_I S_n^+ : (\text{tr}(X))^2 - \mu \text{tr}(X^2) \geq 0, \text{tr}(X) \geq 0\}, \quad (11)$$

defines an $\text{Ad}_{O(n)}$ -invariant cone in $T_I S_n^+ = \{X \in \mathbb{R}^{n \times n} : X^T = X\}$.

Proof $\text{Ad}_{O(n)}$ -invariance is clear since $\text{tr}(X^2) = \text{tr}(QXQ^TQXQ^T)$ and $\text{tr}(X) = \text{tr}(QXQ^T)$ for all $Q \in O(n)$. To prove that (11) is a cone, first note that $0 \in \mathcal{K}(I)$ and for $\lambda > 0$, $X \in \mathcal{K}(I)$, we have $\lambda X \in \mathcal{K}(I)$ since $\text{tr}(\lambda X) = \lambda \text{tr}(X) \geq 0$ and

$$(\text{tr}(\lambda X))^2 - \mu \text{tr}((\lambda X)^2) = \lambda^2[(\text{tr}(X))^2 - \mu \text{tr}(X^2)] \geq 0. \quad (12)$$

To show convexity, let $X_1, X_2 \in \mathcal{K}(I)$. Now $\text{tr}(X_1 + X_2) = \text{tr}(X_1) + \text{tr}(X_2) \geq 0$, and

$$\begin{aligned} (\text{tr}(X_1 + X_2))^2 - \mu \text{tr}((X_1 + X_2)^2) &= [(\text{tr}(X_1))^2 - \mu \text{tr}(X_1^2)] \\ &+ [(\text{tr}(X_2))^2 - \mu \text{tr}(X_2^2)] + 2[\text{tr}(X_1) \text{tr}(X_2) - \mu \text{tr}(X_1 X_2)] \geq 0, \end{aligned} \quad (13)$$

since $\text{tr}(X_1 X_2) \leq (\text{tr}(X_1^2))^{\frac{1}{2}} (\text{tr}(X_2^2))^{\frac{1}{2}} \leq \frac{1}{\sqrt{\mu}} \text{tr}(X_1) \frac{1}{\sqrt{\mu}} \text{tr}(X_2)$, where the first inequality follows by Cauchy-Schwarz. Finally, we need to show that $\mathcal{K}(I)$ is pointed. If $X \in \mathcal{K}(I)$ and $-X \in \mathcal{K}(I)$, then $\text{tr}(-X) = -\text{tr}(X) = 0$. Thus, $(\text{tr}(X))^2 - \mu \text{tr}(X^2) = -\mu \text{tr}(X^2) \geq 0$, which is possible if and only if all of the eigenvalues of X are zero; i.e., if and only if $X = 0$. \square

The parameter μ controls the opening angle of the cone. If $\mu = 0$, then (11) defines the half-space $\text{tr}(X) \geq 0$. As μ increases, the opening angle of the cone becomes smaller and for $\mu = n$ (11) collapses to a ray. For each $\mu \in (0, n)$, the cone $\mathcal{K}_\Lambda = \mathcal{K}_\Lambda^\mu \subset \mathbb{R}^n$ of Proposition 1 is given by

$$\mathcal{K}_\Lambda^\mu = \left\{ \lambda = (\lambda_i) \in \mathbb{R}^n : \left(\sum_{i=1}^n \lambda_i \right)^2 - \mu \sum_{i=1}^n \lambda_i^2 \geq 0, \sum_{i=1}^n \lambda_i \geq 0 \right\}, \quad (14)$$

since $\text{tr}(X) = \sum_{i=1}^n \lambda_i(X)$ and $\text{tr}(X^2) = \sum_{i=1}^n \lambda_i^2(X)$. Indeed \mathcal{K}_Λ^μ is a quadratic cone

$$\mathcal{K}_\Lambda^\mu = \{\lambda \in \mathbb{R}^n : \lambda^T Q_\mu \lambda \geq 0, \mathbf{1}^T \lambda \geq 0\}, \quad (15)$$

where $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^n$, and Q_μ is the $n \times n$ matrix with entries $(Q_\mu)_{ii} = 1 - \mu$ and $(Q_\mu)_{ij} = 1$ for $i \neq j$.

The dual cone C^* of a subset C of a vector space is a very important notion in convex analysis. For a vector space \mathcal{V} endowed with an inner product $\langle \cdot, \cdot \rangle$, the dual cone can be defined as $C^* = \{y \in \mathcal{V} : \langle y, x \rangle \geq 0, \forall x \in C\}$. A cone is said to be self-dual if it coincides with its dual cone. It is well-known that the cone of positive semidefinite matrices is self-dual. The following lemma will be used to characterize the form of the dual cone $(\mathcal{K}_A^\mu)^*$ for each $\mu \in (0, n)$ with respect to the standard inner product on \mathbb{R}^n .

Lemma 1 *The dual cone of the quadratic cone defined by (15) with respect to the standard inner product on \mathbb{R}^n is given by*

$$(\mathcal{K}_A^\mu)^* = \{\lambda \in \mathbb{R}^n : \lambda^T Q_\mu^{-1} \lambda \geq 0, \mathbf{1}^T \lambda \geq 0\}. \quad (16)$$

The inverse matrix Q_μ^{-1} is given by

$$(Q_\mu^{-1})_{ij} = \begin{cases} \frac{\mu - (n-1)}{\mu(n-\mu)} & i = j, \\ \frac{1}{\mu(n-\mu)} & i \neq j. \end{cases} \quad (17)$$

Since $\mu(n - \mu) > 0$ and $\mu - (n - 1) = 1 - \mu^*$ where $\mu^* = n - \mu$, we find that $\lambda^T Q_\mu^{-1} \lambda \geq 0$ if and only if $\lambda^T Q_{\mu^*} \lambda \geq 0$. That is,

$$(\mathcal{K}_A^\mu)^* = \mathcal{K}_A^{n-\mu}. \quad (18)$$

We notice of course from (18) that $\text{Ad}_{O(n)}$ -invariant cones are generally not self-dual. Indeed, for quadratic $\text{Ad}_{O(n)}$ -invariant cones, self-duality is only achieved for $\mu = n/2$.

Now for any fixed $\mu \in (0, n)$, we obtain a unique well-defined affine-invariant cone field given by

$$\mathcal{K}(\Sigma) = \{X \in T_\Sigma S_n^+ : (\text{tr}(\Sigma^{-1} X))^2 - \mu \text{tr}(\Sigma^{-1} X \Sigma^{-1} X) \geq 0, \text{tr}(\Sigma^{-1} X) \geq 0\}. \quad (19)$$

Note that for the value $\mu = 0$, (19) reduces to the affine-invariant half-space field $\{X \in T_\Sigma S_n^+ : \text{tr}(\Sigma^{-1} X) \geq 0\}$. At the other extreme, for $\mu = n$, it is easy to show that the set at I is given by the ray $\{X \in T_I S_n^+ : X = \lambda I, \lambda \geq 0\}$. By affine-invariance, (19) reduces to $\{X \in T_\Sigma S_n^+ : X = \lambda \Sigma, \lambda \geq 0\}$ for $\mu = n$, which describes an affine-invariant field of rays in S_n^+ .

It should be noted that of course not all $\text{Ad}_{O(n)}$ -invariant cones at I are quadratic. Indeed, it is possible to construct polyhedral $\text{Ad}_{O(n)}$ -invariant cones that arise as the intersections of a collection of spectrally defined half-spaces in $T_I S_n^+$. The clearest example of such a construction is the cone of positive semidefinite matrices in $T_I S_n^+$, which of course itself has a spectral characterization $\mathcal{K}(I) = \{X \in T_I S_n^+ : \lambda_i(X) \geq 0, i = 1, \dots, n\}$.

3.2 Affine-invariant pseudo-Riemannian structures on S_n^+

At this point it is instructive to note the following systematic analysis of all affine-invariant pseudo-Riemannian structures on S_n^+ before continuing with our treatment of affine-invariant cone fields. This elegant characterization presents the affine-invariant Riemannian metrics of (6) and the quadratic affine-invariant cone fields of (19) within a unified and rigorous mathematical framework. Recall that a pseudo-Riemannian metric is a generalization of a Riemannian metric in which the metric tensor need not be positive definite, but need only be a non-degenerate, smooth, symmetric bilinear form. The signature of such a metric tensor is defined as the ordered pair consisting of the number of positive and negative eigenvalues of the real and symmetric matrix of the metric tensor with respect to a basis. Note that the signature of a metric tensor is independent of the choice of basis by Sylvester's law of inertia. A metric tensor on a smooth manifold \mathcal{M} is called Lorentzian if its signature is $(1, \dim \mathcal{M} - 1)$.

The irreducible decomposition of \mathfrak{m} under the $\text{Ad}_{O(n)}$ -action is given by $\mathfrak{m} = \mathbb{R}I \oplus \mathfrak{m}_0$, where $\mathfrak{m}_0 := \{X \in \mathfrak{m} : \text{tr } X = 0\}$. According to this decomposition, we have $X = \frac{\text{tr } X}{n}I \oplus \pi(X)$ for any $X \in \mathfrak{m}$, where $\pi(X) := X - \frac{\text{tr } X}{n}I \in \mathfrak{m}_0$. Denote by $\langle X, Y \rangle_{\text{std}}$ the standard inner product $\text{tr}(XY)$ on \mathfrak{m} , and let $\|X\|_{\text{std}}^2 := \langle X, X \rangle_{\text{std}}$ be the corresponding norm. Then we have

$$\text{tr}(X^2) = \|X\|_{\text{std}}^2 = \frac{(\text{tr } X)^2}{n} + \|\pi(X)\|_{\text{std}}^2. \quad (20)$$

Now since \mathfrak{m}_0 is an irreducible $\text{Ad}_{O(n)}$ -module, any $\text{Ad}_{O(n)}$ -invariant quadratic form on \mathfrak{m}_0 is simply a scalar multiple of $\|\cdot\|_{\text{std}}^2$ by Schur's lemma. Therefore, any $\text{Ad}_{O(n)}$ -invariant quadratic form on \mathfrak{m} is of the form

$$Q_{\alpha\beta}(X) := \alpha \frac{(\text{tr } X)^2}{n} + \beta \|\pi(X)\|_{\text{std}}^2, \quad (21)$$

with $\alpha, \beta \in \mathbb{R}$. Clearly, $Q_{\alpha\beta}$ is positive definite if and only if $\alpha > 0$ and $\beta > 0$. Moreover, if $\alpha > 0$ and $\beta < 0$, then $Q_{\alpha\beta}$ is Lorentzian and the set $\{X \in \mathfrak{m} : Q_{\alpha\beta}(X) \geq 0, \text{tr } X \geq 0\}$ defines a pointed cone. Noting that

$$\text{tr}(XY) + \mu \text{tr}(X) \text{tr}(Y) = \left(\mu + \frac{1}{n}\right) \text{tr}(X) \text{tr}(Y) + \langle \pi(X), \pi(Y) \rangle_{\text{std}}, \quad (22)$$

for each $X, Y \in \mathfrak{m}$, we confirm that the metrics in (6) are indeed positive definite if and only if $\mu > -1/n$. Similarly, we find that

$$(\text{tr } X)^2 - \mu \text{tr}(X^2) = \frac{n - \mu}{n} (\text{tr } X)^2 - \mu \|\pi(X)\|_{\text{std}}^2, \quad (23)$$

which is Lorentzian if and only if $0 < \mu < n$. Thus, we see that the affine-invariant pseudo-Riemannian structures on S_n^+ are essentially either Riemannian or Lorentzian, and the quadratic cone fields in (19) are precisely the cone fields defined by the affine-invariant Lorentzian metrics.

3.3 Affine-invariant partial orders on S_n^+

A smooth cone field \mathcal{K} on a manifold \mathcal{M} gives rise to a *conal order* $\prec_{\mathcal{K}}$ on \mathcal{M} , defined by $x \prec_{\mathcal{K}} y$ if there exists a (piecewise) smooth curve $\gamma : [0, 1] \rightarrow \mathcal{M}$ with $\gamma(0) = x$, $\gamma(1) = y$ and $\gamma'(t) \in \mathcal{K}(\gamma(t))$ whenever the derivative exists. The closure $\leq_{\mathcal{K}}$ of this order is again an order and satisfies $x \leq_{\mathcal{K}} y$ if and only if $y \in \{z : x \prec_{\mathcal{K}} z\}$. We say that \mathcal{M} is *globally orderable* if $\leq_{\mathcal{K}}$ is a partial order. Here we will prove that the conal orders induced by affine-invariant cone fields on S_n^+ define partial orders. That is, we will show that the conal orders satisfy the antisymmetry property that $\Sigma_1 \leq_{\mathcal{K}} \Sigma_2$ and $\Sigma_2 \leq_{\mathcal{K}} \Sigma_1$ together imply $\Sigma_1 = \Sigma_2$, for any affine-invariant cone field \mathcal{K} on S_n^+ . In other words, we will prove that there do not exist any non-trivial closed conal curves in S_n^+ . In the following, we will make use of the preimage theorem [3] given below. Recall that given a smooth map $F : \mathcal{M} \rightarrow \mathcal{N}$ between manifolds, we say that a point $y \in \mathcal{N}$ is a *regular value* of F if for all $x \in F^{-1}(y)$ the map $dF|_x : T_x \mathcal{M} \rightarrow T_y \mathcal{N}$ is surjective.

Theorem 1 (The preimage theorem) *Let $F : \mathcal{M} \rightarrow \mathcal{N}$ be a smooth map of manifolds, with $\dim \mathcal{M} = m$ and $\dim \mathcal{N} = n$. If $x \in \mathcal{N}$ is a regular value of F , then $F^{-1}(x)$ is a submanifold of \mathcal{M} of dimension $m - n$. Moreover, the tangent space of $F^{-1}(x)$ at y is equal to $\ker(dF|_y)$.*

Now define $F : S_n^+ \rightarrow \mathbb{R}$ by $F(\Sigma) = \det \Sigma$. By Jacobi's formula, the differential of the determinant takes the form $d(\det)|_{\Sigma} X = \text{tr}(\text{adj}(\Sigma)X)$, where $\text{adj}(\Sigma)$ denotes the adjugate of Σ . That is,

$$dF|_{\Sigma} X = (\det \Sigma) \text{tr}(\Sigma^{-1} X), \quad (24)$$

for all $X \in T_{\Sigma} S_n^+$. Note that for $c > 0$ and any $\Sigma \in F^{-1}(c)$, we have $dF|_{\Sigma} I = c \text{tr}(\Sigma^{-1}) > 0$, which clearly shows that any $c > 0$ is a regular value of F . Hence, $F^{-1}(c)$ is a submanifold of codimension 1 for any choice of $c > 0$. Furthermore, as $\text{im}(F) = \mathbb{R}^+ = \{c \in \mathbb{R} : c > 0\}$, the collection of submanifolds $\{F^{-1}(c)\}_{c>0}$ forms a foliation of S_n^+ . Since $\det \Sigma > 0$ for any $\Sigma \in S_n^+$, (24) implies that $\ker(dF|_{\Sigma}) = \{X \in T_{\Sigma} S_n^+ : \text{tr}(\Sigma^{-1} X) = 0\}$. Thus, the tangent spaces to the submanifolds $\{F^{-1}(c)\}_{c>0}$ are described by the affine-invariant distribution \mathcal{D}_{Σ} of rank $\dim S_n^+ - 1 = n(n+1)/2 - 1$ on S_n^+ defined by $\mathcal{D}_{\Sigma} := \{X \in T_{\Sigma} S_n^+ : \text{tr}(\Sigma^{-1} X) = 0\}$.

Proposition 3 *If $\gamma : [0, 1] \rightarrow S_n^+$ is a non-trivial conal curve with respect to a quadratic affine-invariant cone field \mathcal{K} (19), then*

$$t_2 > t_1 \implies \det(\gamma(t_2)) > \det(\gamma(t_1)), \quad (25)$$

for $t_1, t_2 \in [0, 1]$.

Proof First note that $X \in \mathcal{K}(\Sigma) \setminus \{0\}$ implies that $\text{tr}(\Sigma^{-1} X) > 0$. This follows by noting that if $\text{tr}(\Sigma^{-1} X) = 0$, then $\text{tr}(\Sigma^{-1} X \Sigma^{-1} X) = \text{tr}[(\Sigma^{-1/2} X \Sigma^{-1/2})^2] \leq 0$, which is a contradiction. For simplicity, we assume that γ is a non-trivial

smooth conal curve. The proof for a piecewise smooth curve is similar. We then have $\text{tr}(\gamma(t)^{-1}\gamma'(t)) > 0$, which implies that

$$\frac{d}{dt} \det \gamma(t) = (\det \gamma(t)) \text{tr}(\gamma(t)^{-1}\gamma'(t)) > 0. \quad (26)$$

□

Proposition 3 clearly implies that S_n^+ equipped with any of the cone fields described by (19) does not admit any non-trivial closed conal curves. Indeed, this result holds for all affine-invariant cone fields, not just quadratic ones. To see this, note that the permutation symmetry (10) of Proposition 1, implies that $\text{tr}(\Sigma^{-1}X) \neq 0$ whenever $X \in \mathcal{K}(\Sigma) \setminus \{0\}$. It thus follows by (26) that $\det \circ \gamma : [0, 1] \rightarrow \mathbb{R}^+$ is a strictly monotone function for any non-trivial conal curve γ , which rules out the existence of closed conal curves. We thus arrive at the following theorem.

Theorem 2 *All affine-invariant conal orders on S_n^+ are partial orders.*

At this point it is worth noting a few interesting features of the collection of submanifolds $\{F^{-1}(c)\}_{c>0}$ of S_n^+ . First note that if γ is an *inextensible* conal curve, then by (26) it must intersect each of the submanifolds $F^{-1}(c)$ exactly once. That is, for each $c > 0$, $F^{-1}(c)$ defines a *Cauchy surface* for the causal structure induced by any affine-invariant cone field. We also note the following results which connect these submanifolds to geodesics on S_n^+ with respect to the standard affine-invariant Riemannian metric $ds^2 = \text{tr}[(\Sigma^{-1}d\Sigma)^2]$ on S_n^+ .

Proposition 4 *Endow S_n^+ with the Riemannian structure defined by the standard Riemannian metric $ds^2 = \text{tr}[(\Sigma^{-1}d\Sigma)^2]$. We have the following results.*

- i) *If $\Sigma_1, \Sigma_2 \in S_n^+$ satisfy $\det \Sigma_1 = \det \Sigma_2 = c$, then the geodesic from Σ_1 to Σ_2 lies in $F^{-1}(c)$.*
- ii) *If $X \in T_\Sigma S_n^+$ satisfies $\text{tr}(\Sigma^{-1}X) = 0$, then the geodesic through Σ in the direction of X stays on the submanifold $F^{-1}(\det \Sigma)$.*

Proof i) Let $\Sigma_1, \Sigma_2 \in S_n^+$ satisfy $\det \Sigma_1 = \det \Sigma_2$. The geodesic γ from Σ_1 to Σ_2 is given by

$$\gamma(t) = \Sigma_1^{1/2} \exp \left(t \log \left(\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2} \right) \right) \Sigma_1^{1/2}. \quad (27)$$

Thus, $\det(\gamma(t)) = (\det \Sigma_1) \det(\exp(t \log(\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2})))$. Using the matrix identity $\log(\det A) = \text{tr}(\log A)$, we find that

$$\begin{aligned} \log \left[\det \left(\exp \left(t \log \left(\Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}} \right) \right) \right) \right] &= \text{tr} \left[\log \left(\exp \left(t \log \left(\Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}} \right) \right) \right) \right] \\ &= t \text{tr} \left(\log \left(\Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}} \right) \right) \end{aligned} \quad (28)$$

$$= t \log (\det \Sigma_2 / \det \Sigma_1) = 0. \quad (29)$$

Therefore, $\det(\exp(t \log(\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2})) = 1$, which implies that $\det(\gamma(t)) = \det \Sigma_1$ for all $t \in \mathbb{R}$.

ii) The geodesic γ from Σ in the direction of $X \in T_\Sigma S_n^+$ takes the form $\gamma(t) = \Sigma^{1/2} \exp(t \Sigma^{-1/2} X \Sigma^{-1/2}) \Sigma^{1/2}$. If $\text{tr}(\Sigma^{-1} X) = 0$, then

$$\log(\det(\exp(t \Sigma^{-1/2} X \Sigma^{-1/2}))) = \text{tr}(t \Sigma^{-1/2} X \Sigma^{-1/2}) = t \text{tr}(\Sigma^{-1} X) = 0, \quad (30)$$

which implies that $\det(\gamma(t)) = (\det \Sigma) \det(\exp(t \Sigma^{-1/2} X \Sigma^{-1/2})) = \det \Sigma$ for all $t \in \mathbb{R}$. \square

3.4 Causal semigroups

Define a *wedge* to be a closed and convex subset of a vector space that is also invariant with respect to scaling by positive numbers. Notice in particular that a wedge need not be pointed. Let $\mathcal{M} = G/H$ be a homogeneous space, G a Lie group with group identity element e and Lie algebra \mathfrak{g} , H a closed subgroup with Lie algebra \mathfrak{h} , and $\pi : G \rightarrow \mathcal{M}$ the associated projection map. Assume that the Lie algebra \mathfrak{g} contains a wedge W such that (i) $W \cap -W = \mathfrak{h}$ and (ii) $\text{Ad}(h)W = W$ for all $h \in H$. A wedge W is said to be a *Lie wedge* if $e^{\text{ad } h}W = W$ for all $h \in W \cap -W$. Denoting the left action of G on \mathcal{M} by $\tau_g : \mathcal{M} \rightarrow \mathcal{M}$, we have $\pi \circ \lambda_g = \tau_g \circ \pi$, where λ_g is the left multiplication with g on G . Conditions (i) and (ii) ensure that $d\pi|_g \circ d\lambda_g|_e W$ only depends on $\pi(g)$, so that

$$\mathcal{K}(\pi(g)) = (d\pi|_g \circ d\lambda_g|_e) W, \quad (31)$$

yields a well-defined field of pointed cones on \mathcal{M} that is invariant under the action of G on \mathcal{M} : $d\tau_g|_x \mathcal{K}(x) = \mathcal{K}(\tau_g(x))$. These results can be found in [13]. The set $S = \{g \in G : o \leq_{\mathcal{K}} \tau_g(o)\}$, where $o = \pi(e)$, is a closed semigroup of G referred to as the causal semigroup of $(\mathcal{M}, G, \mathcal{K})$. The following theorem is derived from [18].

Theorem 3 *Let $S = \overline{\langle \exp W \rangle H} \subseteq G$, then $S = \pi^{-1}(\{x \in \mathcal{M} : o \leq_{\mathcal{K}} x\})$ and \mathcal{M} is globally orderable with respect to \mathcal{K} if and only if $W = \mathbf{L}(S)$, where*

$$\mathbf{L}(S) = \{Z \in \mathfrak{g} : \exp(\mathbb{R}^+ Z) \subseteq S\}. \quad (32)$$

The affine-invariant cone fields on $S_n^+ = GL(n)/O(n)$ can be viewed as projections of invariant wedge fields on the Lie group $GL(n)$ in the sense of the above results. Since we have the reductive decomposition $\mathfrak{gl}(n) = \mathfrak{o}(n) \oplus \mathfrak{m}$, it is easy to construct the corresponding wedge field W that satisfies conditions (i) and (ii) for a given affine-invariant cone field \mathcal{K} . We will now use this structure and Theorem 3 to prove the following important result.

Theorem 4 *Let S_n^+ be equipped with an affine-invariant cone field \mathcal{K} and the standard affine-invariant Riemannian metric $ds^2 = \text{tr}[(\Sigma^{-1} d\Sigma)^2]$. For any pair of matrices $\Sigma_1, \Sigma_2 \in S_n^+$, we have $\Sigma_1 \leq_{\mathcal{K}} \Sigma_2$ if and only if the geodesic from Σ_1 to Σ_2 is a conal curve.*

Proof Note that the expression of the geodesic from Σ_1 to Σ_2 given in (27) implies that this theorem is equivalent to

$$\Sigma_1 \leq_{\mathcal{K}} \Sigma_2 \iff \log \left(\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2} \right) \in \mathcal{K}(I). \quad (33)$$

As \mathcal{K} is affine-invariant, $\Sigma_1 \leq_{\mathcal{K}} \Sigma_2$ is equivalent to $I \leq \Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2}$. Thus, it is sufficient to prove that

$$I \leq_{\mathcal{K}} \Sigma \iff \log(\Sigma) \in \mathcal{K}(I), \quad (34)$$

for any $\Sigma \in S_n^+$. We define a wedge W in $\mathfrak{gl}(n)$ by

$$W := \{X + Y : X \in \mathcal{K}(I), Y \in \mathfrak{o}(n)\} \subset \mathfrak{gl}(n) = \mathfrak{m} \oplus \mathfrak{o}(n), \quad (35)$$

where $\mathcal{K}(I)$ is viewed as a subset of $\mathfrak{m} \cong T_I S_n^+$. Note that (35) ensures that W satisfies the properties required of it in Theorem 3. If $I \leq_{\mathcal{K}} \Sigma$, it follows from Theorem 3 that there exists $A \in W$ such that

$$\Sigma = \pi(\exp A) = \tau_{\exp A}(I) = (\exp A)(\exp A)^T. \quad (36)$$

By the polar decomposition theorem of [16], any element $g = \exp A$ of the semigroup $S = \overline{\langle \exp W \rangle O(n)} \subset GL(n)$ admits a unique decomposition as $g = (\exp X)Q$ with $X \in W \cap \mathfrak{m} = \mathcal{K}(I)$ and $Q \in O(n)$. Thus, we have

$$\Sigma = \tau_g(I) = \tau_{\exp X}(I) = \exp 2X, \quad (37)$$

so that $\log \Sigma = 2X \in \mathcal{K}(I)$. \square

Remark 1 Let \mathcal{K} be a quadratic affine-invariant cone field described by (19). Given a pair $\Sigma_1, \Sigma_2 \in S_n^+$, we have by Theorem 4 that $\Sigma_1 \leq_{\mathcal{K}} \Sigma_2$ if and only if $\log \left(\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2} \right) \in \mathcal{K}(I)$, which is equivalent to

$$\begin{cases} \operatorname{tr} \left(\log(\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2}) \right) \geq 0, \\ \left(\operatorname{tr}(\log(\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2})) \right)^2 - \mu \operatorname{tr} \left[(\log(\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2}))^2 \right] \geq 0. \end{cases} \quad (38)$$

Since $\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2}$ and $\Sigma_2 \Sigma_1^{-1}$ have the same spectrum, (38) can be written as

$$\begin{cases} \operatorname{tr}(\log(\Sigma_2 \Sigma_1^{-1})) \geq 0, \\ \left(\operatorname{tr}(\log(\Sigma_2 \Sigma_1^{-1})) \right)^2 - \mu \operatorname{tr}[(\log(\Sigma_2 \Sigma_1^{-1}))^2] \geq 0, \end{cases} \quad (39)$$

which has the virtue of not involving square roots of Σ_1 and Σ_2 . Equation (39) in turn is equivalent to

$$\begin{cases} \sum_i \log \lambda_i \geq 0, \\ (\sum_i \log \lambda_i)^2 - \mu \sum_i (\log \lambda_i)^2 \geq 0, \end{cases} \quad (40)$$

where $\lambda_i = \lambda_i(\Sigma_2 \Sigma_1^{-1})$ ($i = 1, \dots, n$) denote the n real and positive eigenvalues of $\Sigma_2 \Sigma_1^{-1}$. We have thus used invariance to reduce the question of whether a pair of positive definite matrices Σ_1 and Σ_2 are ordered with respect to any of the quadratic affine-invariant cone fields to a pair of inequalities involving the spectrum of $\Sigma_2 \Sigma_1^{-1}$.

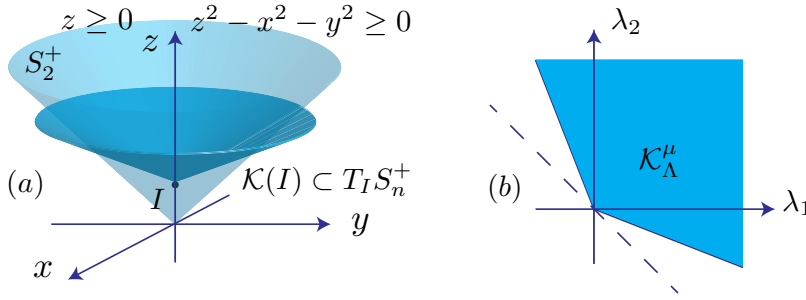


Fig. 1 (a) Identification of S_2^+ with the interior of the closed, convex, pointed cone $K = \{(x, y, z) \in \mathbb{R}^3 : z^2 - x^2 - y^2 \geq 0, z \geq 0\}$ in \mathbb{R}^3 . The $\text{Ad}_{O(n)}$ -invariant cone $\mathcal{K}(I) \subset T_I S_n^+$ at identity is also shown for a choice of $\mu \in (0, 1)$. (b) The corresponding spectral cone $\mathcal{K}_A^\mu \subset \mathbb{R}^2$ which characterizes the cone $\mathcal{K}(I) \subset T_I S_n^+$.

3.5 Visualization of affine-invariant cone fields on S_2^+

It is well-known that the set of positive semidefinite matrices of dimension n forms a cone in the space of symmetric $n \times n$ matrices. Moreover, S_n^+ forms the interior of this cone. A concrete visualization of this identification can be made in the $n = 2$ case, as shown in Figure 1 (a). The set S_2^+ can be identified with the interior of the set $K = \{(x, y, z) \in \mathbb{R}^3 : z^2 - x^2 - y^2 \geq 0, z \geq 0\}$, through the bijection $\phi : S_2^+ \rightarrow \text{int } K$ given by

$$\phi : \begin{pmatrix} a & b \\ b & c \end{pmatrix} \mapsto (x, y, z) = \left(\sqrt{2}b, \frac{1}{\sqrt{2}}(a - c), \frac{1}{\sqrt{2}}(a + c) \right). \quad (41)$$

Inverting ϕ , we find that $a = (z + y)/\sqrt{2}$, $b = x/\sqrt{2}$, $c = (z - y)/\sqrt{2}$. Note that the point $(x, y, z) = (0, 0, \sqrt{2})$ corresponds to the identity matrix $I \in S_2^+$. We seek to arrive at a visual representation of the affine-invariant cone fields generated from the $\text{Ad}_{O(n)}$ -invariant cones (11) for different choices of the parameter μ . The defining inequalities $\text{tr}(X) \geq 0$ and $(\text{tr}(X))^2 - \mu \text{tr}(X^2) \geq 0$ in $T_I S_2^+$ take the forms

$$\delta z \geq 0, \quad \text{and} \quad \left(\frac{2}{\mu} - 1 \right) \delta z^2 - \delta x^2 - \delta y^2 \geq 0, \quad (42)$$

respectively, where $(\delta x, \delta y, \delta z) \in T_{(0,0,\sqrt{2})} K \cong T_I S_2^+$. The corresponding spectral cone $\mathcal{K}_A^\mu \subset \mathbb{R}^2$ is given by

$$\lambda_1 + \lambda_2 \geq 0, \quad \text{and} \quad (\lambda_1 + \lambda_2)^2 - \mu(\lambda_1^2 + \lambda_2^2) \geq 0. \quad (43)$$

See Figure 1 (b) for an illustration of such a cone for a choice of $\mu \in (0, 1)$.

Clearly the translation invariant cone fields generated from this cone are given by the same equations as in (42) for $(\delta x, \delta y, \delta z) \in T_{(x,y,z)} K \cong T_\Sigma S_2^+$,

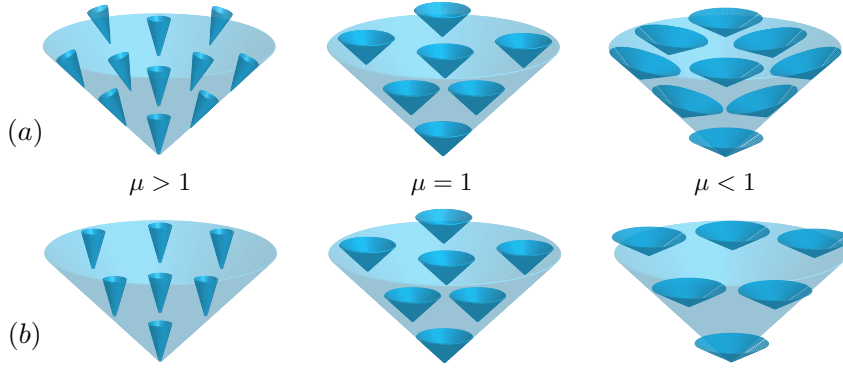


Fig. 2 Cone fields on S_2^+ : (a) Quadratic affine-invariant cone fields for different choices of the parameter $\mu \in (0, 2)$. (b) The corresponding translation-invariant cone fields.

where $\phi(\Sigma) = (x, y, z)$. To obtain the affine-invariant cone fields, note that at $\Sigma = \phi^{-1}(x, y, z) \in S_2^+$, the inequality $\text{tr}(\Sigma^{-1}X) \geq 0$ takes the form

$$\text{tr} \left[\begin{pmatrix} c & -b \\ -b & a \end{pmatrix} \begin{pmatrix} \delta a & \delta b \\ \delta b & \delta c \end{pmatrix} \right] = c \delta a - 2b \delta b + a \delta c \geq 0 \quad (44)$$

$$\iff z \delta z - x \delta x - y \delta y \geq 0. \quad (45)$$

Similarly, the inequality $(\text{tr}(\Sigma^{-1}X))^2 - \mu \text{tr}(\Sigma^{-1}X \Sigma^{-1}X) \geq 0$ is equivalent to

$$2(x \delta x + y \delta y - z \delta z)^2 - \mu [(z^2 + x^2 - y^2) \delta x^2 + (z^2 - x^2 - y^2) \delta y^2 + (x^2 + y^2 + z^2) \delta z^2 + 4xy \delta x \delta y - 4xz \delta x \delta z - 4yz \delta y \delta z] \geq 0, \quad (46)$$

where $(\delta x, \delta y, \delta z) \in T_{(x,y,z)}K \cong T_\Sigma S_2^+$. In the case $\mu = 1$, this reduces to $(\frac{2}{\mu} - 1) \delta z^2 - \delta x^2 - \delta y^2 \geq 0$. Thus, for $\mu = 1$ the quadratic cone field generated by affine-invariance coincides with the corresponding translation-invariant cone field. Generally, however, affine-invariant and translation-invariant cone fields do not agree, as depicted in Figure 2. Each of the distinct cone fields in Figure 2 induces a distinct partial order on S_n^+ .

3.6 The Löwner order

The Löwner order is the partial order \geq_L on S_n^+ defined by

$$A \geq_L B \iff A - B \geq_L O, \quad (47)$$

where the inequality on the right denotes that $A - B$ is positive semidefinite [5]. The definition in (47) is based on translations and the ‘flat’ geometry of S_n^+ . It is clear that the Löwner order is translation invariant in the sense that $A \geq_L B$ implies that $A + C \geq_L B + C$ for all $A, B, C \in S_n^+$. From the perspective of conal orders, the Löwner order is the partial order induced by

the cone field generated by translations of the cone of positive semidefinite matrices at $T_I S_n^+$.

In the previous section, we gave an explicit construction showing that the cone field generated through translations of the cone of positive semidefinite matrices at $T_I S_n^+$ coincides with the cone field generated through affine-invariance in the $n = 2$ case. We will now show that this is a general result which holds for all n . First note that the cone at $T_I S_n^+$ can be expressed as

$$\mathcal{K}(I) = \{X \in T_I S_n^+ : u^T X u \geq 0 \ \forall u \in \mathbb{R}^n, \ u^T X u = 0 \Rightarrow u = 0\}, \quad (48)$$

and the resulting translation-invariant cone field is simply given by

$$\mathcal{K}_T(\Sigma) = \{X \in T_\Sigma S_n^+ : u^T X u \geq 0 \ \forall u \in \mathbb{R}^n, \ u^T X u = 0 \Rightarrow u = 0\}. \quad (49)$$

The corresponding affine-invariant cone field is given by

$$\begin{aligned} \mathcal{K}_A(\Sigma) = \{X \in T_\Sigma S_n^+ : u^T \Sigma^{-1/2} X \Sigma^{-1/2} u \geq 0 \ \forall u \in \mathbb{R}^n, \\ u^T \Sigma^{-1/2} X \Sigma^{-1/2} u = 0 \Rightarrow u = 0\}, \end{aligned} \quad (50)$$

which is seen to be equal to \mathcal{K}_T by introducing the invertible transformation $\bar{u} = \Sigma^{-1/2} u$ in (50). Thus we see that the Löwner order enjoys the special status of being both affine-invariant and translation-invariant, even though its classical definition is based on the ‘flat’ or translational geometry on S_n^+ .

4 Monotone functions on S_n^+

4.1 Differential positivity

Let f be a map of S_n^+ into itself. We say that f is *monotone* with respect to a partial order \geq on S_n^+ if $f(\Sigma_1) \geq f(\Sigma_2)$ whenever $\Sigma_1 \geq \Sigma_2$. Such functions were introduced by Löwner in his seminal paper [17] on operator monotone functions. Since then operator monotone functions have been studied extensively and found applications to many fields including electrical engineering [1], network theory, and quantum information theory [6, 19]. Monotonicity of mappings and dynamical systems with respect to partial orders induced by cone fields have a local geometric characterization in the form of differential positivity [9]. A smooth map $f : S_n^+ \rightarrow S_n^+$ is said to be differentially positive with respect to a cone field \mathcal{K} on S_n^+ if $df|_\Sigma(\delta\Sigma) \in \mathcal{K}(f(\Sigma))$ whenever $\delta\Sigma \in \mathcal{K}(\Sigma)$, where $df|_\Sigma : T_\Sigma S_n^+ \rightarrow T_{f(\Sigma)} S_n^+$ denotes the differential of f at Σ . Assuming that $\geq_{\mathcal{K}}$ is a partial order induced by \mathcal{K} , then f is monotone with respect to $\geq_{\mathcal{K}}$ if and only if it is differentially positive with respect to \mathcal{K} . To see this, recall that $\Sigma_2 \geq_{\mathcal{K}} \Sigma_1$ means that there exists some conal curve $\gamma : [0, 1] \rightarrow S_n^+$ such that $\gamma(0) = \Sigma_1$, $\gamma(1) = \Sigma_2$ and $\gamma'(t) \in \mathcal{K}(\gamma(t))$ for all $t \in (0, 1)$. Now $f \circ \gamma : [0, 1] \rightarrow S_n^+$ is a curve in S_n^+ with $(f \circ \gamma)(0) = f(\Sigma_1)$, $(f \circ \gamma)(1) = f(\Sigma_2)$, and

$$(f \circ \gamma)'(t) = df|_{\gamma(t)} \gamma'(t). \quad (51)$$

Hence, $f \circ \gamma$ is a conal curve joining $f(\Sigma_1)$ to $f(\Sigma_2)$ if and only if $df|_{\gamma(t)}\mathcal{K}(\gamma(t)) \subseteq \mathcal{K}(f(\gamma(t)))$.

4.2 The Generalized Löwner-Heinz Theorem

One of the most fundamental results in operator theory is the Löwner-Heinz theorem [17, 11] stated below.

Theorem 5 (Löwner-Heinz) *If $\Sigma_1 \geq_L \Sigma_2$ in S_n^+ and $r \in [0, 1]$, then*

$$\Sigma_1^r \geq_L \Sigma_2^r. \quad (52)$$

Furthermore, if $n \geq 2$ and $r > 1$, then $\Sigma_1 \geq_L \Sigma_2 \not\Rightarrow \Sigma_1^r \geq_L \Sigma_2^r$.

There are several different proofs of the Löwner-Heinz theorem. See [5, 20, 17, 11], for instance. Most of these proofs are based on analytic methods, such as integral representations from complex analysis. Instead we employ a geometric approach to study monotonicity based on a differential analysis of the system. One of the advantages of such an approach is that it is immediately applicable to all of the conal orders considered in this paper, while providing geometric insight into the behavior of the map under consideration. By using invariant differential positivity with respect to the family of affine-invariant cone fields in (19), we arrive at the following extension to the Löwner-Heinz theorem.

Theorem 6 (Generalized Löwner-Heinz) *For any of the affine-invariant partial orders induced by the quadratic cone fields (19) parametrized by μ , the map $f_r(\Sigma) = \Sigma^r$ is monotone on S_n^+ for any $r \in [0, 1]$.*

This result suggests that the monotonicity of the map $f_r : \Sigma \mapsto \Sigma^r$ for $r \in (0, 1)$ is intimately connected to the affine-invariant geometry of S_n^+ and not its translational geometry. The structure of the proof of Theorem 6 is as follows. We first prove that the map $f_{1/p} : \Sigma \mapsto \Sigma^{1/p}$ is monotone for any $p \in \mathbb{N}$. We then extend this result to maps $f_{q/p} : \Sigma \mapsto \Sigma^{q/p}$ for rational numbers $q/p \in \mathbb{Q} \cap (0, 1)$, before arriving at the full result via a density argument. We prove monotonicity by establishing differential positivity in each case. To prove the monotonicity of $f_{1/p} : \Sigma \mapsto \Sigma^{1/p}$, $p \in \mathbb{N}$, we only need the following lemma [24].

Lemma 2 *If A and B are Hermitian $n \times n$ matrices, then*

$$\mathrm{tr}[(AB)^{2m}] \leq \mathrm{tr}[A^{2m}B^{2m}], \quad m \in \mathbb{N}. \quad (53)$$

The proof of the theorem for rational exponents is based on a simple observation whose proof nonetheless requires a few technical steps that are based on Proposition 5, which itself relies on Lemma 3 established in [7, 10].

Lemma 3 Let F, G be real-valued functions on some domain $D \subseteq \mathbb{R}$ and Σ, X be Hermitian matrices, such that the spectrum of Σ is contained in D . If (F, G) is an antimonotone pair so that $(F(a) - F(b))(G(a) - G(b)) \leq 0$ for all $a, b \in D$, then

$$\operatorname{tr} [F(\Sigma)XG(\Sigma)X] \geq \operatorname{tr} [F(\Sigma)G(\Sigma)X^2]. \quad (54)$$

Proposition 5 If $\Sigma \in S_n^+$ and X is a Hermitian matrix, then

$$\operatorname{tr} (\Sigma^{-2-k} X \Sigma^k X) \geq \operatorname{tr} (\Sigma^{-1-k} X \Sigma^{-1+k} X), \quad (55)$$

for integers $k \geq 0$.

Proof Define $F, G : (0, \infty) \rightarrow \mathbb{R}$ by $F(x) := x^{-1-2k}$ and $G(x) := x$, and note that $(F(a) - F(b))(G(a) - G(b)) \leq 0$ for all $a, b > 0$. Let $\Sigma \in S_n^+$ and X be a Hermitian matrix. Then, we have

$$\begin{aligned} \operatorname{tr} (\Sigma^{-2-k} X \Sigma^k X) &= \operatorname{tr} \left[\Sigma^{-1-2k} \left(\Sigma^{-\frac{1+k}{2}} X \Sigma^{-\frac{1+k}{2}} \right) \Sigma \left(\Sigma^{-\frac{1+k}{2}} X \Sigma^{-\frac{1+k}{2}} \right) \right] \\ &\geq \operatorname{tr} \left[\Sigma^{-2k} \left(\Sigma^{-\frac{1+k}{2}} X \Sigma^{-\frac{1+k}{2}} \right) \left(\Sigma^{-\frac{1+k}{2}} X \Sigma^{-\frac{1+k}{2}} \right) \right] \end{aligned} \quad (56)$$

$$= \operatorname{tr} (\Sigma^{-1-k} X \Sigma^{-1+k} X), \quad (57)$$

following an application of Lemma 3 with the Hermitian matrix replaced by $\Sigma^{(-1+k)/2} X \Sigma^{(-1+k)/2}$. \square

Proof (of Theorem 6) : The differential $df_{1/p}|_\Sigma : T_\Sigma S_n^+ \rightarrow T_{f_{1/p}(\Sigma)} S_n^+$ of $f_{1/p}$ satisfies the generalized Sylvester equation

$$\sum_{j=0}^{p-1} (\Sigma^{1/p})^{p-1-j} (df_{1/p}|_\Sigma X) (\Sigma^{1/p})^j = X, \quad (58)$$

for every $X \in T_\Sigma S_n^+$. Thus,

$$\sum_{j=0}^{p-1} (\Sigma^{1/p})^{p-1-j-\frac{1}{2}p} (df_{1/p}|_\Sigma X) (\Sigma^{1/p})^{j-\frac{1}{2}p} = \Sigma^{-1/2} X \Sigma^{-1/2}. \quad (59)$$

Taking the trace of (59) yields

$$\operatorname{tr} \left(\sum_{j=0}^{p-1} (\Sigma^{1/p})^{\frac{1}{2}p-1-j} (df_{1/p}|_\Sigma X) (\Sigma^{1/p})^{j-\frac{1}{2}p} \right) = \operatorname{tr} (\Sigma^{-1/2} X \Sigma^{-1/2}) \quad (60)$$

$$\implies \operatorname{tr} \left(\sum_{j=0}^{p-1} \Sigma^{-1/p} (df_{1/p}|_\Sigma X) \right) = \operatorname{tr} (\Sigma^{-1} X). \quad (61)$$

That is, $p \operatorname{tr}((f_{1/p}(\Sigma))^{-1}(df_{1/p}|_{\Sigma}X)) = \operatorname{tr}(\Sigma^{-1}X)$, for all $X \in T_{\Sigma}S_n^+$. Now taking the trace of the square of (59), we obtain

$$\operatorname{tr} \left(\sum_{i,j=0}^{p-1} (\Sigma^{1/p})^{i-j-1} (df_{1/p}|_{\Sigma}X) (\Sigma^{1/p})^{j-i-1} (df_{1/p}|_{\Sigma}X) \right) = \operatorname{tr}(\Sigma^{-1}X \Sigma^{-1}X). \quad (62)$$

The left-hand side of (62) can be rewritten as

$$\sum_{i,j=0}^{p-1} \operatorname{tr} \left[\left((\Sigma^{1/p})^{\frac{j-i-1}{2}} (df_{1/p}|_{\Sigma}X) (\Sigma^{1/p})^{\frac{j-i-1}{2}} \right)^2 \left((\Sigma^{1/p})^{i-j} \right)^2 \right] \quad (63)$$

$$\geq \sum_{i,j=0}^{p-1} \operatorname{tr} \left[\left((\Sigma^{1/p})^{\frac{j-i-1}{2}} (df_{1/p}|_{\Sigma}X) (\Sigma^{1/p})^{\frac{j-i-1}{2}} (\Sigma^{1/p})^{i-j} \right)^2 \right] \quad (64)$$

$$= \sum_{i,j=0}^{p-1} \operatorname{tr} \left[\Sigma^{-1/p} (df_{1/p}|_{\Sigma}X) \Sigma^{-1/p} (df_{1/p}|_{\Sigma}X) \right] \quad (65)$$

$$= p^2 \operatorname{tr} \left[(f_{1/p}(\Sigma))^{-1} (df_{1/p}|_{\Sigma}X) (f_{1/p}(\Sigma))^{-1} (df_{1/p}|_{\Sigma}X) \right], \quad (66)$$

where the inequality follows from an application of Lemma 2. Thus,

$$\operatorname{tr} \left[((f_{1/p}(\Sigma))^{-1} (df_{1/p}|_{\Sigma}X))^2 \right] \leq \frac{1}{p^2} \operatorname{tr}(\Sigma^{-1}X \Sigma^{-1}X). \quad (67)$$

Combined with (61), this implies that

$$\begin{aligned} & [\operatorname{tr}((f_{1/p}(\Sigma))^{-1}(df_{1/p}|_{\Sigma}X))]^2 - \mu \operatorname{tr} \left[((f_{1/p}(\Sigma))^{-1}(df_{1/p}|_{\Sigma}X))^2 \right] \\ & \geq \frac{1}{p^2} ([\operatorname{tr}(\Sigma^{-1}X)]^2 - \mu \operatorname{tr}(\Sigma^{-1}X \Sigma^{-1}X)) \geq 0, \end{aligned} \quad (68)$$

for all $X \in \mathcal{K}(\Sigma)$. That is, $(df_{1/p}|_{\Sigma})\mathcal{K}(\Sigma) \subseteq \mathcal{K}(f_{1/p}(\Sigma))$ for any choice of μ .

This result can be extended to all rational powers $q/p \in \mathbb{Q} \cap [0, 1]$ by combining two observations. First, since the inverse of the p -th root matrix function $f_{1/p}$ is the p -th power function $f_p : \Sigma \mapsto \Sigma^p$ and $f_{1/p}$ contracts the invariant cone field \mathcal{K} , f_p must expand \mathcal{K} . Second, this expansion is greater for larger p . That is, for positive integers $p_1 \leq p_2$,

$$(d\tau_{\Sigma^{-1/2p_1}}|_{\Sigma^{p_1}} \circ df_{p_1}|_{\Sigma})\mathcal{K}(\Sigma) \subseteq (d\tau_{\Sigma^{-1/2p_2}}|_{\Sigma^{p_2}} \circ df_{p_2}|_{\Sigma})\mathcal{K}(\Sigma). \quad (69)$$

Thus, the map $f_{q/p} = f_q \circ f_{1/p}$ is differentially positive, since the contraction of the cone field by $f_{1/p}$ will dominate the expansion of the cone field by f_q for $p \geq q$. Note that the contractions and expansions referred to here need not be strict for the argument to hold. To prove (69), it is sufficient to show that the map f_{p+1} expands the cone field at least as much as f_p for any $p \in \mathbb{N}$. This is done by showing that

$$df_p|_{\Sigma}X \in \partial\mathcal{K}(\Sigma^p) \implies df_{p+1}|_{\Sigma}X \notin \operatorname{int}\mathcal{K}(\Sigma^{p+1}), \quad (70)$$

for any $\Sigma \in S_n^+$ and $X \in T_\Sigma S_n^+$, where $\partial\mathcal{K}(\Sigma^p)$ denotes the boundary of $\mathcal{K}(\Sigma^p)$. Note that $df_p|_\Sigma X \in \partial\mathcal{K}(\Sigma^p)$ implies that $X \in \mathcal{K}(\Sigma)$, since f_p expands \mathcal{K} . The implication in (70) shows that the expansion of the cone field by f_{p+1} is at least as great as that of f_p by linearity of the differential maps. Using $\text{tr}(f_p(\Sigma)^{-1}df_p|_\Sigma X) = p \text{tr}(\Sigma^{-1}X)$, we see that $df_p|_\Sigma X \in \partial\mathcal{K}(\Sigma^p)$ is equivalent to

$$p^2 \text{tr}(\Sigma^{-1}X)^2 = \mu \sum_{i,j=0}^{p-1} \text{tr}(\Sigma^{-1+i-j}X \Sigma^{-1+j-i}X). \quad (71)$$

Assuming (71), we have

$$\begin{aligned} & \left[\text{tr}((f_{p+1}(\Sigma))^{-1}(df_{p+1}|_\Sigma X)) \right]^2 - \mu \text{tr} \left[((f_{p+1}(\Sigma))^{-1}(df_{p+1}|_\Sigma X))^2 \right] \\ &= (p+1)^2 \text{tr}(\Sigma^{-1}X)^2 - \mu \sum_{i,j=0}^p \text{tr}(\Sigma^{-1+i-j}X \Sigma^{-1+j-i}X) \\ &= \frac{\mu(p+1)^2}{p^2} \sum_{i,j=0}^{p-1} \text{tr}(\Sigma^{-1+i-j}X \Sigma^{-1+j-i}X) - \mu \sum_{i,j=0}^p \text{tr}(\Sigma^{-1+i-j}X \Sigma^{-1+j-i}X), \end{aligned} \quad (72)$$

where the last equation follows from substitution using (71). Using the simplification $\sum_{i,j=0}^{p-1} \text{tr}(\Sigma^{-1+i-j}X \Sigma^{-1+j-i}X) = \sum_{k=0}^{p-1} \alpha_k \text{tr}(\Sigma^{-k-1}X \Sigma^{k-1}X)$, where $\alpha_0 = p$ and $\alpha_k = 2(p-k)$ for $k \geq 1$, (72) reduces to

$$\begin{aligned} & \mu \left[\left(p \frac{(p+1)^2}{p^2} - (p+1) \right) \text{tr}(\Sigma^{-1}X \Sigma^{-1}X) \right. \\ & \quad \left. + \frac{(p+1)^2}{p^2} \sum_{k=1}^{p-1} 2(p-k) \text{tr}(\Sigma^{-1-k}X \Sigma^{-1+k}X) \right. \\ & \quad \left. - \sum_{k=1}^p 2(p+1-k) \text{tr}(\Sigma^{-1-k}X \Sigma^{-1+k}X) \right] \end{aligned} \quad (73)$$

$$\begin{aligned} &= \mu \left[\frac{p+1}{p} \text{tr}(\Sigma^{-1}X \Sigma^{-1}X) \right. \\ & \quad \left. + \sum_{k=1}^{p-1} \beta_k \text{tr}(\Sigma^{-1-k}X \Sigma^{-1+k}X) - 2 \text{tr}(\Sigma^{-1-p}X \Sigma^{-1+p}X) \right], \end{aligned} \quad (74)$$

where

$$\beta_k = 2 \frac{(p+1)^2(p-k)}{p^2} - 2(p+1-k). \quad (75)$$

We find that $\beta_k \geq 0$ if and only if $k \leq l := \lfloor p/2 \rfloor$, where $\lfloor \cdot \rfloor$ identifies the integer part of its argument. Thus, through repeated applications of Proposition 5,

we see that (74) is less than or equal to

$$\begin{aligned} & \mu \left(\frac{p+1}{p} + \sum_{k=1}^l \beta_k \right) \operatorname{tr} (\Sigma^{-1-l} X \Sigma^{-1+l} X) - \mu \left(2 + \sum_{k=l+1}^{p-1} |\beta_k| \right) \operatorname{tr} (\Sigma^{-2-l} X \Sigma^l X) \\ &= \mu \left(2 + \frac{(p-l-1)(l+2pl-p)}{p^2} \right) [\operatorname{tr} (\Sigma^{-1-l} X \Sigma^{-1+l} X) - \operatorname{tr} (\Sigma^{-2-l} X \Sigma^l X)], \end{aligned} \quad (76)$$

which is nonpositive by a final application of Proposition 5. This completes the proof of (70).

Finally, we extend the result to all real exponents $r \in [0, 1]$. Assume for a contradiction that there exists some $r \in (0, 1)$ and $\Sigma_1, \Sigma_2 \in S_n^+$ such that $\Sigma_1 \geq \Sigma_2$ and $\Sigma_1^r < \Sigma_2^r$. Define $E = \{x \in (0, 1) : \Sigma_1^x < \Sigma_2^x\}$ and note that $E \neq \emptyset$ since $r \in E$. As E is an open set in \mathbb{R} , there exists some $s \in \mathbb{Q} \cap E$ so that $\Sigma_1^s < \Sigma_2^s$, which is a contradiction. Therefore, f_r is monotone for all $r \in [0, 1]$ with respect to any of the affine-invariant orders parametrized by μ . \square

Remark 2 The geometric insight provided by differential positivity clarifies the duality between the monotonicity of the function $f_r : \Sigma \mapsto \Sigma^r$ for $0 < r < 1$ and its non-monotonicity for $r > 1$, which may seem somewhat mysterious otherwise. Specifically, since the inverse of the function f_r is given by $f_{1/r}$, we see that if f_r contracts affine-invariant cone fields for $r \in (0, 1)$ at every point, then $f_{1/r}$ must expand the same cone fields. Indeed, if the contraction of \mathcal{K} by f_r is strict at some $\Sigma \in S_n^+$, then $f_{1/r}$ cannot be differentially positive with respect to \mathcal{K} and so is not monotone with respect to $\leq_{\mathcal{K}}$. See Figure 3. To show that this is indeed the case for any of the affine-invariant cone fields (19), we note that at any $\Sigma \in S_n^+$, $X_{\Sigma} = \Sigma \in T_{\Sigma} S_n^+$ lies in the interior of $\mathcal{K}(\Sigma)$, since $(\operatorname{tr}(\Sigma^{-1} X_{\Sigma}))^2 - \mu \operatorname{tr}(\Sigma^{-1} X_{\Sigma} \Sigma^{-1} X_{\Sigma}) = n^2 - \mu n > 0$ and $\operatorname{tr}(\Sigma^{-1} X_{\Sigma}) = \operatorname{tr}(I) = n > 0$ for $\mu \in (0, n)$. Let $\Sigma = \operatorname{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$ be any diagonal matrix in S_n^+ with $\sigma_1 > \sigma_2$. As $X_{\Sigma} = \Sigma \in \operatorname{int} \mathcal{K}(\Sigma)$, there exists some $\delta > 0$ such that

$$X = (x_{ij}) = \begin{pmatrix} \sigma_1 & \delta & & \\ \delta & \sigma_2 & & \\ & & \sigma_3 & \\ & & & \ddots \\ & & & & \sigma_n \end{pmatrix} \quad (77)$$

lies on the boundary of $\partial \mathcal{K}(\Sigma)$. Specifically, we find that

$$(\operatorname{tr}(\Sigma^{-1} X))^2 - \mu \operatorname{tr}(\Sigma^{-1} X \Sigma^{-1} X) \quad (78)$$

$$= \left(\sum_i \frac{x_{ii}}{\sigma_i} \right)^2 - \mu \left(\sum_i \frac{x_{ii}^2}{\sigma_i^2} + \frac{2}{\sigma_1 \sigma_2} \delta^2 \right) = n^2 - \mu \left(n + \frac{2}{\sigma_1 \sigma_2} \delta^2 \right) \quad (79)$$

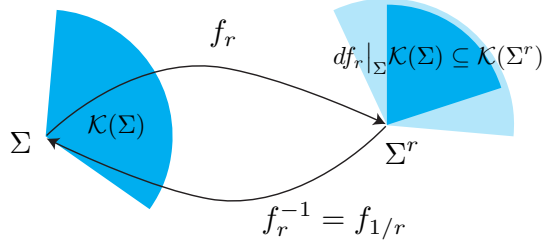


Fig. 3 Contraction of affine-invariant cone fields by $f_r : \Sigma \mapsto \Sigma^r$ for $0 < r < 1$ corresponds to expansion of affine-invariant cone fields by the inverse map $f_r^{-1} = f_{1/r} : \Sigma \mapsto \Sigma^{1/r}$.

vanishes when

$$\delta^2 = \frac{n(n-\mu)\sigma_1\sigma_2}{2\mu}. \quad (80)$$

Now for this choice of X , the inequality (55) of Proposition 5 with $k = 0$ becomes strict as

$$\text{tr}(\Sigma^{-1}X\Sigma^{-1}X) = n + \frac{2}{\sigma_1\sigma_2}\delta^2 < n + \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)\delta^2 = \text{tr}(\Sigma^{-2}X^2), \quad (81)$$

since $(1/\sigma_1 - 1/\sigma_2)^2 > 0$. As this inequality is used to derive (76), which is used to prove (69), it follows that the contraction of \mathcal{K} by f_r is strict at some $\Sigma \in S_n^+$ for $r \in (0, 1)$. Therefore, f_r cannot be monotone with respect to $\leq_{\mathcal{K}}$ for $r > 1$.

4.3 Matrix inversion

Consider the matrix inversion map $f(\Sigma) = \Sigma^{-1}$. The differential $df|_{\Sigma} : T_{\Sigma}S_n^+ \rightarrow T_{\Sigma^{-1}}S_n^+$ of f is given by

$$df|_{\Sigma}X = -\Sigma^{-1}X\Sigma^{-1}. \quad (82)$$

To show this, it is sufficient to consider the geodesic from Σ in the direction $X \in T_{\Sigma}S_n^+$ given by

$$\gamma(t) = \Sigma^{1/2} \exp(t\Sigma^{-1/2}X\Sigma^{-1/2})\Sigma^{1/2}, \quad (83)$$

and note that $(f \circ \gamma)(t) = \Sigma^{-1/2} \exp(-t\Sigma^{-1/2}X\Sigma^{-1/2})\Sigma^{-1/2}$ so that

$$(f \circ \gamma)'(0) = \Sigma^{-1/2}(-\Sigma^{-1/2}X\Sigma^{-1/2})e^{-t\Sigma^{-1/2}X\Sigma^{-1/2}}\Sigma^{-1/2}\big|_{t=0} = -\Sigma^{-1}X\Sigma^{-1}. \quad (84)$$

Thus, $\text{tr}(\Sigma(df|_{\Sigma}X)) = -\text{tr}(\Sigma^{-1}X)$ and $\text{tr}[(\Sigma df|_{\Sigma}X)^2] = \text{tr}(\Sigma^{-1}X\Sigma^{-1}X)$. Therefore, noting the conditions in (19), it is clear that $\Sigma \mapsto \Sigma^{-1}$ reverses the ordering of positive definite matrices for any of the affine-invariant orders since

$$\text{tr}((f(\Sigma))^{-1}(df|_{\Sigma}X)) = -\text{tr}(\Sigma^{-1}X). \quad (85)$$

That is,

$$\Sigma_1 \geq_{\mathcal{K}} \Sigma_2 \implies \Sigma_2^{-1} \geq_{\mathcal{K}} \Sigma_1^{-1}, \quad (86)$$

for any of the affine-invariant cone fields \mathcal{K} in (19).

4.4 Scaling and congruence transformations

Consider the function $S_\lambda : S_n^+ \rightarrow S_n^+$ defined by $S_\lambda(\Sigma) = \lambda\Sigma$, where $\lambda > 0$ is a scalar. The differential $dS_\lambda|_\Sigma : T_\Sigma S_n^+ \rightarrow T_{\lambda\Sigma} S_n^+$ is given by $dS_\lambda|_\Sigma X = \lambda X$. Substituting into the formula for the family of quadratic affine-invariant cones (19), we find that

$$\begin{aligned} & \left[\text{tr} \left(S_\lambda(\Sigma)^{-1} (dS_\lambda|_\Sigma X) \right) \right]^2 - \mu \text{tr} \left(S_\lambda(\Sigma)^{-1} (dS_\lambda|_\Sigma X) S_\lambda(\Sigma)^{-1} (dS_\lambda|_\Sigma X) \right) \\ &= \left[\text{tr} \left(\frac{1}{\lambda} \Sigma^{-1} \lambda X \right) \right]^2 - \mu \text{tr} \left(\frac{1}{\lambda} \Sigma^{-1} \lambda X \right)^2 = [\text{tr}(\Sigma^{-1} X)]^2 - \mu \text{tr}(\Sigma^{-1} X)^2 \geq 0 \end{aligned} \quad (87)$$

for any $X \in \mathcal{K}(\Sigma)$. Thus, S_λ is differentially positive and so preserves the affine-invariant orders induced by any of the cone fields (19). This is of course a special case of a more general result about congruence transformations $\tau_A(\Sigma) = A\Sigma A^T$, where $A \in GL(n)$. Congruence transformations can be thought of as generalizations of scaling transformations on S_n^+ . The preservation of affine-invariant orders by congruence transformations follows by construction. If $\Sigma_1 \leq_{\mathcal{K}} \Sigma_2$ for some partial order induced by an affine-invariant cone field \mathcal{K} , then there exists a conal curve γ from Σ_1 to Σ_2 . It follows from the definition of affine-invariant cone fields that congruence transformations map conal curves to conal curves in S_n^+ . That is, $\tau_A(\gamma(t))$ is a conal curve joining $\tau_A(\Sigma_1)$ to $\tau_A(\Sigma_2)$.

4.5 Translations

It is important to note that translations do not generally preserve an affine-invariant order unless the associated affine-invariant cone field happens to also be translation invariant.

Proposition 6 *Let $\leq_{\mathcal{K}}$ denote the partial order induced by an affine-invariant cone field \mathcal{K} on S_n^+ . If \mathcal{K} is not translation invariant, then there exists a translation $T_C : S_n^+ \rightarrow S_n^+$, $T_C(\Sigma) = \Sigma + C$ that does not preserve $\leq_{\mathcal{K}}$.*

Proof If \mathcal{K} is not translation invariant, then there exist $\Sigma_1, \Sigma_2 \in S_n^+$ such that $dT_{(\Sigma_2 - \Sigma_1)}|_{\Sigma_1} \mathcal{K}(\Sigma_1) \neq \mathcal{K}(\Sigma_2)$, where $T_{(\Sigma_2 - \Sigma_1)}(\Sigma) = \Sigma + (\Sigma_2 - \Sigma_1)$. Thus there exists some $\delta\Sigma$ in the cone at either Σ_1 or Σ_2 that cannot be identified with an element of the cone at the other point under translation. Without loss of generality, assume that $\delta\Sigma \in \mathcal{K}(\Sigma_1)$ and $dT_{(\Sigma_2 - \Sigma_1)}|_{\Sigma_1}(\delta\Sigma) \notin \mathcal{K}(\Sigma_2)$. For an affine-invariant cone field \mathcal{K} , we have

$$\mathcal{K}(\lambda\Sigma) = d\tau_{\lambda^{1/2}I}|_\Sigma \mathcal{K}(\Sigma) = dS_\lambda|_\Sigma \mathcal{K}(\Sigma) = \lambda\mathcal{K}(\Sigma) = \mathcal{K}(\Sigma) \quad (88)$$

for any $\lambda > 0$ and $\Sigma \in S_n^+$. That is, the cone field is translationally invariant along each ray $\gamma(t) = t\Sigma$, $t > 0$. Thus, we can identify $\mathcal{K}(\Sigma_2)$ through translation with any cone $\mathcal{K}(\lambda\Sigma_2)$ where $\lambda > 0$. It follows that $dT_{(\lambda\Sigma_2 - \Sigma_1)}|_{\Sigma_1}(\delta\Sigma) \notin \mathcal{K}(\lambda\Sigma_2)$ for any $\lambda > 0$. For sufficiently large $\lambda > 0$, $C := \lambda\Sigma_2 - \Sigma_1$ is a positive definite matrix. Therefore, $T_C : S_n^+ \rightarrow S_n^+$ is not differentially positive with respect to \mathcal{K} and hence is not monotone with respect to $\leq_{\mathcal{K}}$. \square

5 Invariant half-spaces

5.1 An affine-invariant half-space preorder

The $\text{Ad}_{O(n)}$ -invariant condition $\text{tr}(X) \geq 0$ on $T_I S_n^+$ in (11) picks out a pointed cone from the double cone defined by the non-negativity of the quadratic form $(\text{tr}(X))^2 - \mu \text{tr}(X^2)$. Indeed, $\text{tr}(X) \geq 0$ defines a half-space in $T_I S_n^+$ bounded by the hyperplane $\text{tr}(X) = 0$ in $T_I S_n^+$. The affine-invariant extension of this hyperplane to all of S_n^+ yields a distribution of rank $\dim S_n^+ - 1 = n(n+1)/2 - 1$ on S_n^+ given by $\text{tr}(\Sigma^{-1/2} X \Sigma^{-1/2}) = \text{tr}(\Sigma^{-1} X) = 0$ for $X \in T_{\Sigma} S_n^+$. The corresponding affine-invariant half-space field \mathcal{H}_{Σ} on the tangent bundle $T S_n^+$ simply takes the form

$$\mathcal{H}_{\Sigma} = \{X \in T_{\Sigma} S_n^+ : \text{tr}(\Sigma^{-1} X) \geq 0\}. \quad (89)$$

A half-space field of this form induces a partial preorder $\preceq_{\mathcal{H}}$ on S_n^+ . That is, a binary relation that is reflexive and transitive. The antisymmetry condition required for a preorder to be a partial order does not hold since \mathcal{H}_{Σ} is not a pointed cone. Nonetheless, one can ask whether any two given matrices $\Sigma_1, \Sigma_2 \in S_n^+$ satisfy $\Sigma_1 \preceq_{\mathcal{H}} \Sigma_2$, or if a given function on S_n^+ is monotone with respect to the preorder induced by (89). The monotonicity of a function with respect to a preorder still gives geometric insight into the effects of the function on the space on which it acts and the discrete-time dynamics defined by its iterations.

To illustrate this we return to a puzzling aspect concerning the monotonicity of the function $f_r(x) = x^r$ on the real line for $r > 0$ and its analogue result for positive semidefinite matrices. Namely, that the map f_r is monotone on S_n^+ with respect to an affine-invariant partial order if $r \in [0, 1]$ but is not monotone on S_n^+ for $r > 1$. We will show that the monotonicity on the real line for $r > 0$ is inherited in the matrix function setting in the form of a one-dimensional monotonicity expressed as the preservation of the affine-invariant half-space preorder for any $r > 0$.

Proposition 7 *The function $f_r : \Sigma \mapsto \Sigma^r$ is monotone on S_n^+ with respect to the affine-invariant half-space preorder $\preceq_{\mathcal{H}}$ for any $r > 0$.*

Proof Let $p, q \in \mathbb{N}$ be positive integers. The map $f_{q/p} : \Sigma \mapsto \Sigma^{q/p}$ can be written as the composition $f_{1/p} \circ f_q$ with differential

$$df_{q/p}|_{\Sigma} = df_{1/p}|_{f_q(\Sigma)} \circ df_q|_{\Sigma}. \quad (90)$$

Now since $df_q|_\Sigma$ is given by

$$df_q|_\Sigma X = \sum_{j=0}^{q-1} \Sigma^{q-1-j} X \Sigma^j, \quad (X \in T_\Sigma S_n^+) \quad (91)$$

and $df_{1/p}|_\Sigma$ is the unique solution of the generalized Sylvester equation (58), the differential $df_{q/p}|_\Sigma$ in (90) must satisfy

$$\sum_{i=0}^{p-1} (\Sigma^{q/p})^{p-1-i} (df_{q/p}|_\Sigma X) (\Sigma^{q/p})^i = \sum_{j=0}^{q-1} \Sigma^{q-1-j} X \Sigma^j. \quad (92)$$

Multiplying both sides of this equation by Σ^{-q} and taking the trace of the resulting equation yields

$$\text{tr} \left(\sum_{i=0}^{p-1} (\Sigma^{q/p})^{-1-i} (df_{q/p}|_\Sigma X) (\Sigma^{q/p})^i \right) = \text{tr} \left(\sum_{j=0}^{q-1} \Sigma^{-1-j} X \Sigma^j \right) \quad (93)$$

$$\implies \text{tr} \left(\sum_{i=0}^{p-1} \Sigma^{-q/p} (df_{q/p}|_\Sigma X) \right) = \text{tr} \left(\sum_{j=0}^{q-1} \Sigma^{-1} X \right) \quad (94)$$

$$\implies p \text{tr} \left(\Sigma^{-q/p} (df_{q/p}|_\Sigma X) \right) = q \text{tr} (\Sigma^{-1} X). \quad (95)$$

That is, $\text{tr} ((f_{q/p}(\Sigma))^{-1} df_{q/p}|_\Sigma X) = \frac{q}{p} \text{tr} (\Sigma^{-1} X)$ for all $X \in T_\Sigma S_n^+$. A standard argument based on the density of positive rational numbers in the positive real line \mathbb{R}_+ gives

$$\text{tr} ((f_r(\Sigma))^{-1} df_r|_\Sigma X) = r \text{tr} (\Sigma^{-1} X) \quad (96)$$

for any real $r > 0$. Therefore, we clearly have the implication

$$X \in \mathcal{H}_\Sigma \implies df_r|_\Sigma X \in \mathcal{H}_{f_r(\Sigma)} \quad (97)$$

for all $X \in T_\Sigma S_n^+$, which is precisely the local characterization of the monotonicity of f_r with respect to the preorder induced by \mathcal{H}_Σ . \square

This result further highlights the natural connection between affine-invariance of causal structures on S_n^+ and monotonicity of the matrix power functions $f_r(\Sigma) = \Sigma^r$. In particular, f_r is generally not monotone with respect to a preorder induced by a half-space field that is translation-invariant.

It should be noted that although the above proof has the virtue of being self-contained, Proposition 7 can also be proven using results from Section 3.3. Specifically, it should be clear from the material from that section that $\Sigma_1 \preceq_{\mathcal{H}} \Sigma_2$ if and only if $\det \Sigma_1 \preceq_{\mathcal{H}} \det \Sigma_2$, whence $f_r : \Sigma \mapsto \Sigma^r$ preserves $\preceq_{\mathcal{H}}$ precisely when

$$\det \Sigma_1 \leq \det \Sigma_2 \implies \det f_r(\Sigma_1) \leq \det f_r(\Sigma_2). \quad (98)$$

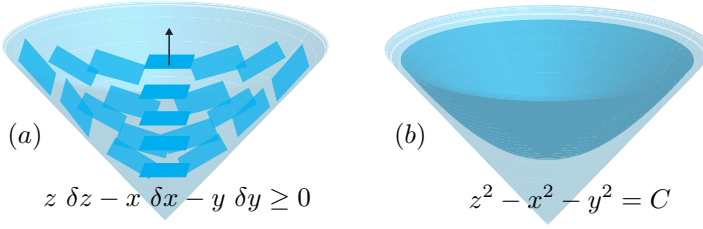


Fig. 4 (a) An illustration of the affine-invariant hyperplanes $\partial\mathcal{H}$ corresponding to $\text{tr}(\Sigma^{-1}X) = 0$ against the backdrop of the cone $K = \{(x, y, z) \in \mathbb{R}^3 : z^2 - x^2 - y^2 \geq 0, z \geq 0\}$ identified with S_2^+ . (b) The distributions integrate to give a family of hyperboloids of revolution parametrized by $C > 0$. The limiting case $C = 0$ yields the boundary of the cone K .

Since $\det f_r(\Sigma) = \det \Sigma^r = r(\det \Sigma)$, this is clearly the case for any $r > 0$.

It is instructive to return to the $n = 2$ case to obtain a visualization of the rank 2 distribution $\mathcal{D}_\Sigma = \partial\mathcal{H}$ that defines the affine-invariant preorder induced by \mathcal{H}_Σ . As noted in Section 3.5, the set S_2^+ can be identified with the interior of the quadratic cone K in \mathbb{R}^3 given by $z^2 - x^2 - y^2 \geq 0, z \geq 0$ via a bijection $\phi : \Sigma \mapsto (x, y, z)$. At $\Sigma = \phi^{-1}(x, y, z) \in S_2^+$, the inequality $\text{tr}(\Sigma^{-1}X) \geq 0$ takes the form $z\delta z - x\delta x - y\delta y \geq 0$, where $(\delta x, \delta y, \delta z) \in T_{(x,y,z)}K$ as shown in (45). The distribution $\partial\mathcal{H}$ that consists of the hyperplanes which form the boundary of the half-space field \mathcal{H}_Σ are given by $z\delta z - x\delta x - y\delta y = 0$. This distribution is clearly integrable with integral submanifolds of the form $z^2 - x^2 - y^2 = C$, where $C \geq 0$ is a constant for each of the integral submanifolds, which form hyperboloids of revolution as shown in Figure 4. As expected, these surfaces coincide with the submanifolds of constant determinant predicted in Section 3.3.

5.2 The Toda and QR flows

The *Toda* flow is a well-know Hamiltonian dynamical system on the space of real symmetric matrices of fixed dimension n , which can be expressed in the Lax pair form

$$\dot{X}(t) = [X, \pi_s(X)] = X\pi_s(X) - \pi_s(X)X, \quad (99)$$

where $\pi_s(X)$ is the skew-symmetric matrix $\pi_s(X) = X_{ij}$ if $i > j$, $\pi_s(X) = 0$ if $i = j$, and $\pi_s(X) = -X_{ji}$ if $i < j$. The QR-flow is a related dynamical system on S_n^+ that has close connections to the QR algorithm and is given by

$$\dot{\Sigma}(t) = [\Sigma, \pi_s(\log \Sigma)]. \quad (100)$$

The Lax pair formulations of the Toda and QR-flows show that these flows are *isospectral*. That is, the eigenvalues of $X(t)$ and $\Sigma(t)$ are independent of t . Isospectral flows clearly preserve all translation invariant orders that possess spectral characterizations.

In [15], the following theorem is established for the projected Toda and QR flows. The projected flows refer to projections of the flows to the $r \times r$ upper left corner principal submatrices of $X(t)$ and $\Sigma(t)$, i.e., the flows of $X_r(t) = E_r^T X(t) E_r$ and $\Sigma_r(t) = E_r^T \Sigma(t) E_r$, where $E_r^T = [I_r \ 0]$.

Theorem 7 *For $1 \leq r \leq n$ and any symmetric matrix $X(0)$ and symmetric positive definite matrix $\Sigma(0)$, the ordered eigenvalues of the projected Toda flow orbit $X_r(t) = E_r^T X(t) E_r$ and the projected QR flow orbit $\Sigma_r(t) = E_r^T \Sigma(t) E_r$ are nondecreasing functions of t .*

Corollary 1 *Let $f(x)$ be any nondecreasing real-valued function and $\alpha > 0$. Then $F(t) = \text{tr}(f(E_r^T X(t) E_r))$ and $G(t) = \text{tr}(f(E_r^T \Sigma(t)^\alpha E_r))$ are nondecreasing functions of t for $t \in \mathbb{R}$.*

The geometric interpretation of the above corollary is that the generalized projected Toda and QR flows, $f(X_r(t))$ and $f(\Sigma_r(t))$, respectively, preserve the half-space preorder induced by the translation invariant half-space $\text{tr}(X) \geq 0$. This is clear by noting that if $X(0), \hat{X}(0)$ are symmetric matrices such that $\text{tr}(X(0) - \hat{X}(0)) \geq 0$, then

$$\text{tr}(f(X_r(t)) - f(\hat{X}_r(t))) \geq \text{tr}(X(0) - \hat{X}(0)) \geq 0, \quad \forall t > 0, \quad (101)$$

and similarly for the generalized projected QR flow.

6 Matrix means

Notions of means and averaging operations on matrices are of great interest in matrix analysis and operator theory with numerous applications to fields such as radar data processing, medical imaging, statistics and machine learning. Adapting basic properties of means on the positive real line to the setting of positive definite matrices, we may define a matrix mean to be a continuous map $M : S_n^+ \times S_n^+ \rightarrow S_n^+$ that satisfies the following properties

1. $M(\Sigma_1, \Sigma_2) = M(\Sigma_2, \Sigma_1)$
2. $\Sigma_1 \leq \Sigma_2 \implies \Sigma_1 \leq M(\Sigma_1, \Sigma_2) \leq \Sigma_2$
3. $M(A^T \Sigma_1 A, A^T \Sigma_2 A) = A^T M(\Sigma_1, \Sigma_2) A$, for all $A \in GL(n)$.
4. $M(\Sigma_1, \Sigma_2)$ is monotone in Σ_1 and Σ_2 .

In the existing literature on matrix means, the partial order \leq in the above definition refers to the Löwner order \leq_L . It is a nontrivial question whether a given map $M : S_n^+ \times S_n^+ \rightarrow S_n^+$ defines a matrix mean with respect to any of the new partial orders considered in this paper. A particularly important matrix mean that has been the subject of considerable interest in recent years is the geometric mean $M(\Sigma_1, \Sigma_2) = \Sigma_1 \# \Sigma_2$ defined by

$$\Sigma_1 \# \Sigma_2 = \Sigma_1^{1/2} \left(\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2} \right)^{1/2} \Sigma_1^{1/2}. \quad (102)$$

The following theorem shows that the geometric mean and affine-invariant orders on S_n^+ are intimately connected.

Theorem 8 *The geometric mean $\#$ (102) defines a matrix mean for any affine-invariant order \leq on S_n^+ .*

Proof The geometric mean $\Sigma_1 \# \Sigma_2$ of two points $\Sigma_1, \Sigma_2 \in S_n^+$ is the midpoint of the geodesic joining Σ_1 and Σ_2 in S_n^+ endowed with the standard Riemannian metric $ds^2 = \text{tr}[(\Sigma^{-1}d\Sigma)^2]$ [5]. This geometric interpretation immediately implies $\Sigma_1 \# \Sigma_2 = \Sigma_2 \# \Sigma_1$. Furthermore, given any affine-invariant order $\leq_{\mathcal{K}}$ induced by an affine-invariant cone field \mathcal{K} and a pair of matrices satisfying $\Sigma_1 \leq_{\mathcal{K}} \Sigma_2$, the geodesic $\gamma : [0, 1] \rightarrow S_n^+$ from Σ_1 to Σ_2 is a conal curve by Theorem 4. Hence, the midpoint $\Sigma_1 \# \Sigma_2$ of γ clearly satisfies $\Sigma_1 \leq_{\mathcal{K}} \Sigma_1 \# \Sigma_2 \leq_{\mathcal{K}} \Sigma_2$. Since congruence transformations are isometries, for any $A \in GL(n)$ the geodesic connecting $A^T \Sigma_1 A$ to $A^T \Sigma_2 A$ is given by $\tilde{\gamma}(t) = A^T \gamma(t) A$. Thus, $(A^T \Sigma_1 A) \# (A^T \Sigma_2 A) = A^T (\Sigma_1 \# \Sigma_2) A$. Finally, for fixed $\Sigma_1 \in S_n^+$, the function $F(\Sigma) = \Sigma_1 \# \Sigma$ is monotone with respect to any affine-invariant order since congruence transformations preserve affine-invariant orders and the function $\Sigma \mapsto \Sigma^{1/2}$ is monotone for any affine-invariant order. By symmetry, $\#$ is also monotone with respect to its first argument. That is, the four conditions that define a matrix mean are all satisfied by the geometric mean for any choice of affine-invariant order. \square

7 Conclusion

The choice of partial order is a key part of studying monotonicity of functions that is often taken for granted. Invariant cone fields provide a geometric approach to systematically construct ‘natural’ orders by connecting the geometry of the state space to the search for orders. Coupled with differential positivity, invariant cone fields provide an insightful and powerful method for studying monotonicity, as shown in the case of S_n^+ . Future work can focus on exploring the applications of the new partial orders presented in this paper to the study of dynamical systems and convergence analysis of algorithms defined on matrices. It may also be fruitful to explore the implications of this work in convexity theory. New notions of partial orders mean new notions of convexity. In this context it may be natural to consider the concept of geodesic convexity on S_n^+ with respect to the Riemannian structure on S_n^+ , as well as the usual notion of convexity on sets of matrices that is based on translational geometry.

Acknowledgements We should like to thank the anonymous referees whose reviews resulted in significant improvements to the quality of this paper. We are particularly grateful to the reviewer who suggested the elegant characterization of affine-invariant pseudo-Riemannian structures presented in Section 3.2 and important clarifications in Section 3.4.

References

1. Anderson WN, Trapp GE (1976) A class of monotone operator functions related to electrical network theory. *Linear Algebra and its Applications* 15(1):53 – 67
2. Ando T (1979) Concavity of certain maps on positive definite matrices and applications to hadamard products. *Linear Algebra and its Applications* 26:203 – 241
3. Banyaga A, Hurtubise D (2013) *Lectures on Morse homology*, vol 29. Springer Science & Business Media
4. Barbaresco F (2008) Innovative tools for radar signal processing based on Cartan's geometry of SPD matrices & information geometry. In: *Proceedings of the IEEE International Radar Conference*, Rome, Italy, IEEE
5. Bhatia R (2007) *Positive Definite Matrices*. Princeton University Press
6. Bhatia R (2013) *Matrix analysis*, vol 169. Springer Science & Business Media
7. Bourin JC (1999) Some inequalities for norms on matrices and operators. *Linear Algebra and its Applications* 292(1):139 – 154
8. Burbea J, Rao C (1982) Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. *Journal of Multivariate Analysis* 12(4):575 – 596
9. Forni F, Sepulchre R (2016) Differentially positive systems. *IEEE Transactions on Automatic Control* 61(2):346–359
10. Fujii JI (2005) A trace inequality arising from quantum information theory. *Linear Algebra and its Applications* 400:141 – 146
11. Heinz E (1951) Beiträge zur strungstheorie der spektralzerlegung. *Mathematische Annalen* 123:415–438
12. Hilgert J, Ólafsson G (1997) *Causal symmetric spaces: Geometry and harmonic analysis*, vol 18. Elsevier
13. Hilgert J, Hofmann KH, Lawson J (1989) *Lie groups, convex cones, and semi-groups*. Oxford University Press
14. Kubo F, Ando T (1979) Means of positive linear operators. *Mathematische Annalen* 246:205–224
15. Lagarias JC (1991) Monotonicity properties of the toda flow, the qr-flow, and subspace iteration. *SIAM J Matrix Anal Appl* 12(3):449–462
16. Lawson JD (1991) Polar and Ol'shanski decompositions. In: *Seminar Sophus Lie*, vol 1, pp 163–173
17. Löwner K (1934) Über monotone matrixfunktionen. *Mathematische Zeitschrift* 38:177–216
18. Neeb KH (1991) Conal orders on homogeneous spaces. *Inventiones mathematicae* 104(1):467–496
19. Nielsen MA, Chuang I (2002) *Quantum computation and quantum information*
20. Pedersen GK (1972) Some operator monotone functions. *Proceedings of the American Mathematical Society* 36(1):309–310
21. Pennec X (2006) *Statistical Computing on Manifolds for Computational Anatomy*. Habilitation à diriger des recherches, Université Nice Sophia Antipolis, URL <https://tel.archives-ouvertes.fr/tel-00633163>
22. Segal IE (1976) *Mathematical cosmology and extragalactic astronomy*, vol 68. Academic Press
23. Skovgaard LT (1984) A Riemannian geometry of the multivariate normal model. *Scandinavian Journal of Statistics* 11(4):211–223
24. Yang Z, Feng X (2002) A note on the trace inequality for products of hermitian matrix power. *JIPAM Journal of Inequalities in Pure & Applied Mathematics* [electronic only] 3(5)